

## **Assigning 2010 Block Codes to 2011 and 2013 American Housing Survey Records**

Issued April 23, 2015

### **Introduction**

In 2014, HUD requested that the Bureau determine the 2010 Census Block for each of the 295,162 AHS housing units in the combined 2011 and 2013 sample. The purpose of the exercise was so Census could use the 2010 Census Block to attach additional geographic variables to the AHS Internal Use File (IUF), thereby making the geographic variables available in the AHS Table Creator, as well as for researchers accessing the data through Census Research Data Centers. Examples of the geography HUD intends to add to the AHS IUF include: 2010 Census Tracts and Block Groups, 2010 Urban Area; 2013 Metro Area and Micro Area; 2010 Rural-Urban Commuting Areas (RUCA); 2013 Rural-Urban Continuum Codes (RUCC); and 2010 Public Use Micro Areas (PUMA).

### **2010 Census Block Codes**

Each one of the geographies mentioned above is a different combination of 2010 Census Blocks (hereinafter referred to as 2010 Block(s)). As such, in order to attach these geographies to the AHS housing units, we first needed to assign the 2010 Block to each of the 295,162 AHS housing units from the combined 2011/2013 sample. Our first attempt to assign 2010 Blocks used the MAF since MAFIDs had previously been assigned to most AHS housing units. The MAF contains the 2010 Block for most MAFIDs. After the MAFID matching process, 14,973 AHS housing units were still missing the 2010 Block. We then submitted the addresses of these 14,973 AHS housing units through the Census geocoder. After the geocoder, 6,148 AHS housing units were still missing the 2010 Block.

### **Imputation Methodology**

An AHS geography imputation program was developed based on a similar process developed for the American Community Survey (ACS). The imputation process uses housing units on the MAF to develop probabilities that a AHS housing unit located in a particular state, county, and zip code combination will be located within a certain 2010 Block.

The first pass through the program matched the 9-digit zip code of the AHS housing unit with 9-digit zip codes found in the MAF. For example, suppose the 9-digit zip code of an AHS housing unit existed on 100 housing units in the MAF. Further suppose that 90 of those MAF housing units were in Block 1 and 10 were in Block 2. The AHS housing unit would then have had a 90% chance of being assigned to Block 1 and a 10% chance of being assigned to Block 2.

For AHS housing units with a 9-digit zip code that did not appear in the MAF, the above imputation process is repeated using only the first 8 digits of the 9-digit zip code. This imputation process was repeated for 7-, 6-, and 5-digit zip codes.

For AHS housing units with a 5-digit zip code that did not appear in the MAF, the above imputation process was conducted using only the state and county codes. Since all AHS housing units had state and county codes, the imputation process was certain to succeed at county imputation level.

Table 1 below shows the distribution of the imputation level applied for the 6,148 AHS housing units. Over 90% of the AHS housing units were imputed using the state, county, and 5-digit zip code. This was expected, because the 5-digit zip code was typically the only zip code present on the address fields for most housing units.

Table 1. Type of Imputation for AHS Housing units with Missing 2010 Block

Imputation Level	Number of AHS Housing Units	Percent of Total
County	195	3.17%
5-Digit Zip	5,570	90.60%
6-Digit Zip	23	0.37%
7-Digit Zip	52	0.85%
8-Digit Zip	158	2.57%
9-Digit Zip	150	2.44%
Total	6,148	100%

### Analysis of Imputation Methodology Accuracy

In order to evaluate the accuracy of the imputation methodology, the same imputation process was applied to the 280,189 AHS housing units with a known 2010 Block, based on an exact match in the MAF. Table 2 shows the results of the imputation process. The results show a similar trend as above - the majority of the housing units were imputed using the state, county, and 5-digit zip code. However, a larger number of housing units were imputed with the 9-digit zip code. This is most likely due to the 4-digit additional zip field being more likely to be present for AHS housing units that had enough address information to be assigned 2010 Block through the MAF or geocoding match process.

Table 2. Type of Imputation for AHS Housing Units with Known 2010 Block

Imputation Level	Number of AHS Housing Units	Percent of Total
County	1,181	0.42%
5-Digit Zip	200,721	71.64%
6-Digit Zip	659	0.24%
7-Digit Zip	600	0.21%
8-Digit Zip	2,795	1.00%
9-Digit Zip	74,233	26.49%
Total	280,189	100%

Table 3 below shows how the known 2010 Block (Tract) compares to the imputed 2010 Block (Tract), for housing units that were matched by county or zip code or 9-digit zip code.

The results show that, for 2010 Block imputations based on the 9-digit zip code, 91.3% of housing units had an imputed 2010 Block that was an exact match to the actual 2010 Block, and 98.7% of housing units had an exact match between the actual 2010 Tract and the imputed 2010 Tract. However, for 2010 Block imputations based on the 5-digit zip code, only 1.2 % of housing units had an imputed 2010 Block that was an exact match to the actual 2010 Block, while 21% had an exact match between the actual 2010 Tract and the imputed 2010 Tract.

Table 3. Differences Between Known and Imputed 2010 Block, by Imputation Level

<b>Imputation Level</b>	<b>Number of AHS Housing Units</b>	<b>2010 Block Match</b>	<b>2010 Tract Match</b>
County	1,181	0.1%	6.6%
5-Digit Zip	200,721	1.2%	21.0%
6-Digit-Zip	659	0.6%	17.5%
7-Digit-Zip	600	15.0%	58.2%
8-Digit-Zip	2,795	18.2%	76.8%
9-Digit Zip	74,233	91.3%	98.7%
Total	280,189	25.3%	42.1%

## Conclusion

Fortunately, the Census Bureau is able to provide an exact 2010 Block for 98% of AHS housing units, based on either an address match in the MAF or through the Census geocoder. However, for the 2% of AHS housing units for which an exact address match cannot be made, the imputation methodology does not provide much help for correctly identifying the 2010 Block. This is because the majority of those 2% of AHS housing units have only a 5-digit zip code (as opposed to a 9-digit zip code). The 5-digit zip code is simply not enough information to make an accurate 2010 Block assignment (1.2% of the time) or 2010 Tract assignment (21.0% percent of the time).